Women in Machine Learning workshop - 2016

Title: Stochastic Function Norm Regularization of Deep Networks

Date of presentation: 05/12/2016

Author: Amal Rannen Triki
Emails: arannen@esat.kuleuven.be, amal.rannen@yonsei.ac.kr
Institutions: KU Leuven, Yonsei University

Supervised by: Matthew B. Blaschko
Email: matthew.blaschko@esat.kuleuven.be
Institution: KU Leuven

Keywords:

- Deep neural networks
- Regularization
- Regularized network training
- L2 function norm
- Weighted norm
- Slice sampling
- Stochastic gradient descent
- Convolutional networks

Abstract:

Although deep neural networks (DNNs) have dramatically improved image classification systems, they require massive labeled databases to train accurate systems. Indeed, DNNs have a complicated mathematical structure. They are highly non-convex, and classical approaches to training on smaller datasets, such as L_2 regularization, cannot be straightforwardly applied. This has precluded their application in important domains where large datasets are not available, such as medical image analysis. This lack of data is due to the high cost of image acquisition, frequently requiring specialized equipment, and the need of expensive, highly qualified experts to annotate the collected samples. Furthermore, many important diseases are rare, meaning that few training examples are available even when image acquisition is inexpensive.

In this work, we propose a new regularization method for DNN training. The state-of-the-art methods are based mainly on weight decay and DropOut. These methods yield impressive performance when a large training set is available, but their performance decreases when applied to only few hundreds of data samples. Moreover, they deal with the complexity of the function represented by the neural network indirectly. Our idea is to study the feasibility of directly regularizing with the L_2 function norm.

The computation of the function norm is not easily accessible, mainly because of the problem high dimension. Thus, we consider using weighted norms. Rather than integrating the network function with respect to the Lebesgue measure, we propose to use two probability measures. The first measure is the data distribution. To obtain an unbiased estimate of this weighted norm, we operate in two steps: (i) at each step of training, the network is applied to a batch of images that are not in the training

set; (ii) we compute the average of the l2 norm of these outputs. The images used for this estimation do not need expert annotation. The second distribution is proportional to $||f(x)||^2$. The main difference with the first method is that in the step (i), the network is applied to samples drawn from this distribution, avoiding the need to acquire new images. The regularization samples are drawn using slice sampling.

For both of the methods, we provide an algorithm integrating them in backpropagation. Moreover, the convergence of the stochastic gradient descent applied to these new objectives is proved. We finally show that the proposed methods applied to convolutional networks outperform the state-of-the-art methods in the low sample regime on benchmark datasets (MNIST and CIFAR10). The obtained results demonstrate clear improvement. The methods have also been applied to classify Optical Coherence Tomography images of breast tissue to detect the presence of cancer in an intraoperative environment. The obtained results show very good qualitative performance, and are currently being evaluated for possible clinical application in a major university hospital.

The proposed methods are very promising in the context of small sample regimes with data lying in a low dimensional manifold.

A extended version of this work is available as an arXiv preprint at https://arxiv.org/pdf/1605.09085.pdf, and an open source implementation can be found at https://github.com/AmalRT/DNN_Reg.